

Available online at www.sciencedirect.com**SciVerse ScienceDirect**

Procedia - Social and Behavioral Sciences 54 (2012) 169 – 178

Procedia
Social and Behavioral Sciences

EWGT 2012

15th meeting of the EURO Working Group on Transportation

Inclusion of the latent personality variable in multinomial logit models using the 16pf psychometric test

J.E. Córdoba^{a,*}, G.P. Jaramillo^b^a*School of Civil Engineering, University National of Colombia, Medellin, Colombia*^b*School of System Engineering, University National of Colombia, Medellin, Colombia*

Abstract

Travel demand models typically use modal attributes and socioeconomic characteristics as explanatory variables. It has been established that attitudes and perceptions as well as individual psychological variables influence a user's behavior. In this study, the latent personality variable was included in the estimation of a hybrid discrete choice model to incorporate the effects of subjective factors. The latent personality variable was assessed with the 16PF psychometric test, which has been widely used by researchers worldwide. The paper analyzes the results of applying this model to a sample of employees and university professors and proposes a way in which the psychometric tests can be used in hybrid discrete choice models. Our results show that hybrid models that include latent psychological variables are superior to traditional models that ignore the effects of user's behavior.

© 2012 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of the Program Committee

Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Latent psychological variable; Discrete choice model; Multinomial logit model; Personality; Psychometric test

1. Introduction

Travel demand models typically use modal attributes and socioeconomic characteristics as explanatory variables of the choice modal. It has been established that attitudes and perceptions influence a user's behavior in the choice modal, and in the last decade, hybrid discrete choice models, which include a latent variable model and a discrete choice model, have been developed that can account for attitudes and perceptions as well as modal attributes and socioeconomic characteristics (Ben-Akiva et al., 2002). Furthermore, the inclusion of latent

* Corresponding author. Tel.: 574-425-5189; fax: 574-425-5150.
E-mail address: jecordob@unal.edu.co

variables improves the fit of these choice models as seen in Raveau et al. (2010). Traditional discrete choice models have been enriched with the construction of latent variables by Ashok et al. (2002), McFadden (1986), Morikawa and Sasaki (1998), Morikawa et al. (2002), Pendleton and Shonkwiler (2001), Vredin et al. (2005) and Yáñez et al. (2010), but the research carried out by Walker (2001) released the first complete methodology for the inclusion of latent variables in discrete choice models. However, there have been no studies thus far that use a personality variable, such as a psychological aspect, for modal choice. Furthermore, psychometric tests have not been used to measure the latent indicators of latent variables.

To estimate the hybrid discrete choice model with the latent personality variable, a sequential approach was used in which the latent variable was first constructed previous to the estimation by the multiple indicator multiple cause (MIMIC) model and then included in the discrete choice model as a regular variable (Vredin et al., 2005).

The main objectives of this research were to estimate a hybrid discrete choice model to include psychological issues, such as personality, and psychometric tests to measure latent variables. The results of an application of this model to a population of employees and professors in a university of Medellín (Colombia) were reviewed, and a methodology for the use of psychometric tests in the hybrid discrete choice models was proposed. Our results show that hybrid models that include psychological latent variables are superior to traditional models that ignore the effects of user's behavior.

2. Theoretical framework

A model is the simplified representation of reality with a mathematical framework because it takes the most representative variables of a system and evaluates their impacts on the system by testing several alternatives. In this section, econometric discrete choice models and latent variables models are presented.

2.1. Econometric discrete choice models

There is a microeconomic analysis of consumer behavior based on the fundamental assumption that the rational consumer will always choose the combination of alternatives more useful for him among those belonging to the set of feasible alternatives. This analysis includes psychological variables such as personality. The set of feasible alternatives for the set of all combinations that consumers can choose, $p = (p_1, p_2, \dots, p_k)$, is the vector of prices of all goods X and the income I available to consumer q . The set of possible combinations is given by Eq. 1, (Williams and Ortúzar, 1982).

$$A(q) = \{x \in X: p \cdot x \leq I\} \quad (1)$$

Thus, the problem facing the consumer can be expressed as Eq. 2.

$$\text{Max } U(x) \text{ s.t. } p \cdot x \leq I \quad x \in X \quad (2)$$

The random utility theory (Domencich and McFadden, 1975) was used, for estimate discrete choice model which states that individuals belonging to a certain homogeneous population act rationally and have perfect information.

2.2. Multinomial Logit Model (MNL)

This model is obtained by assuming that the error terms are independently and identically distributed (i.i.d.) with a Gumbel distribution. This distribution it is also known as extreme value, type I extreme value, and Weibull, with zero mean and variance σ^2 . Therefore, the terms are uncorrelated and have the same variance for each alternative and each individual (Domencich and McFadden, 1975).

2.3. *Modeling with latent variables*

Latent variables are abstract variables representing the subjective elements in the choice conduct; they cannot be measured directly, so they are expressed by only the individual through latent indicators. The methodology developed by Ben-Akiva et al. (2002) is used here to incorporate latent variables as explanatory factors in discrete choice models.

2.4. *Theory of the Eysenck personality*

A study on the theory of personality (Cattell y Eysenck, 1967) through the factorial model seeks intermediate variables that explain differences in behavior in similar situations, as well as the consequences of such behavior. The theory defines personality as the sum of the behavior patterns and potential of the organism, both of which are determined by heredity and the social environment in which the organism originated and developed through the functional interaction of four main factors: (a) cognitive sector (intelligence), (b) conative sector (character), (c) affective sector (temperament), and (d) somatic sector (constitution). Using the theory of Eysenck, Cattell et al. (1970) made extensive use of the factorial analysis method and isolated 16 personality factors, which he brought together in a psychometric test called 16PF. The most relevant aspects of this test are presented below.

2.5. *16PF Psychometric Testing*

This test consists of 187 questions evaluating 16 factors, each of which is measured in decatypes (a score of 1 to 10). These factors are described below (Cattell et al., 1970).

Intellectual area (B)

Personal area (A - E - H - I - M - N-O)

Emotional area (C - G - Q3 - Q1 - Q4)

Social area (F - L - N - Q2)

Factor A (reserved–open) measures the individual's gregarious nature, defined as the degree to which the person seeks to establish contact with other people because they find satisfying and rewarding relationships through them. **Factor B** (concrete thinking–abstract thinking) measures intelligence based on the predominance of abstract or concrete thinking, where abstract thinking is characteristic of a person of higher intelligence and concrete thinking is an indicator of lower intelligence. **Factor C** (emotional instability–emotional stability) is related to the emotional stability of the person and the way in which he adapts to his environment. This factor specifically determines the strength of the ego.

Factor E (submissive–dominant) measures the degree of control that the person tends to hold in their relationships with other human beings and is determined in terms of whether the person is dominant or submissive. **Factor F** (prudent-impulsive) is related to the level of enthusiasm evident in social contexts. **Factor G** (carefree-scrupulous) measures the internalization of moral values. It structurally explores the superego. **Factor H** (shy-spontaneous) measures the reactivity of the nervous system based on the parasympathetic or sympathetic dominance trends of the person.

Factor I (rational-emotional) is used to measure the prevalence of either feeling or rational thought in making decisions for behaving in everyday life. **Factor L** (trusting-suspicious) explores the social identity of the

individual and specifically measures the degree to which the person is identified or linked to the human race in general. **Factor M** (practical-dreamer) is based on the observation that humans can perceive things in two ways. The first way is to receive feed from direct contact between the senses and the environment. The other way is composed mostly of a subliminal connection of thoughts and speculations through which information is organized. **Factor N** (single-sly) is related to social masks and describes the extent to which people are hidden, showing only those features that generate the answers you want from others. **Factor O** (safe-unsafe) explores the self-esteem of those trends based on experience, guilt, or insecurities. This factor is not intended to categorize people by high and low self-esteem, as the level at the time of the test may be a transient because it is influenced by recent events.

Finally, **Factor Q1** (traditionalist-innovative) explores the psychological orientation toward change. **Factor Q2** (dependence on the group self-sufficient) measures the degree of dependence on the person. **Factor Q3** (uninhibited-controlled) explores the efforts of the individual to maintain congruence between their ideal and real selves, molding according to standards established and approved by society. **Factor Q4** (calm-stressed) measures the unpleasant sensations that tend to accompany the excitation of the autonomic nervous system, commonly known as stress.

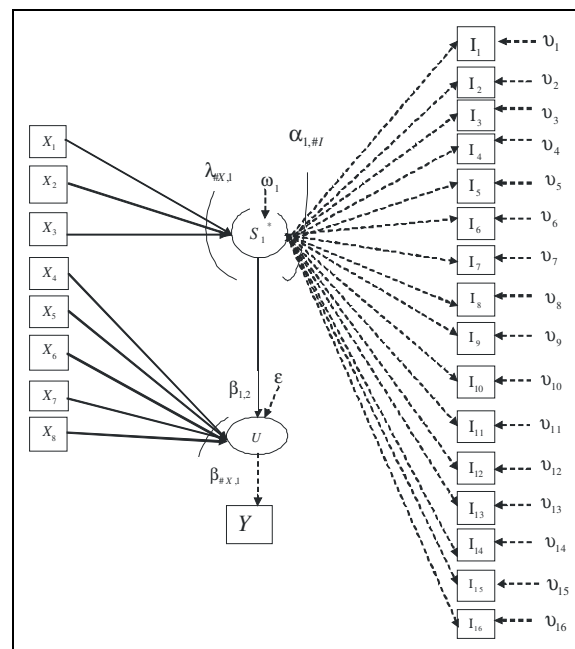


Figure 1 Integrated choice model with latent variable Personality

3. Proposed model

This research is about building a more realistic choice behavior model that incorporates latent constructs such as personality. The responses to the questions to the 16PF personality psychometric test are used as indicators of the latent psychological aspects (see Figure 1).

There are 3 exogenous explanatory variables and 16 indicators for the latent personality variable from the 16PF psychometric test assessment. The model equations are given by Eq. 3.

Structural equation model:

$$S_{ln}^* = X_n \lambda_l + \omega_{ln} \quad l=1 \quad \omega_n \sim N(0, \Sigma_{\omega} \text{diagonal}) \quad (3)$$

$$S_{ln}^* = X_n \lambda_1 + \omega_{ln} \quad \omega_n \sim N(0, \Sigma_{\omega} \text{diagonal})$$

$$X_n = X_1, X_2, X_3$$

X_1 = Sex, X_2 = Age, and X_3 = Education

Where:

S_{ln}^* = Latent personality of individual n

X_n = Observed variables, including socioeconomic characteristics of individual n and attributes of alternative i and individual n .

λ_1 = Unknown personality parameter

ω_n = Error term in the personality equation the resulting utility equation is given by Eq. 4.

$$U_n = X_n \beta_1 + S_{ln}^* \beta_2 + \varepsilon_n \quad \varepsilon_n \sim N(0,1) \quad (4)$$

$$X_n = X_4, X_5, X_6, X_7, X_8$$

With:

X_4 = Cost, X_5 = Travel time, X_6 = Cost by Income, X_7 = Cost by Sex, and X_8 = Walking time

U_n = Utility vector

Where

S_{ln}^* = Latent personality of the individual n

β_1 = Unknown parameter of the utility that has to do with the exogenous variables

β_2 = Unknown parameter of the utility that has to do with the latent variable of personality

ε_n = Error term in the utility equation

The 16 equations (one per indicator) for measuring the latent variables through the indicators are as follows (Eq. 5):

$$I_m = S_{ln}^* \alpha_r + v_m \quad r = 1, 2, 3, \dots, 16 \quad v_m \sim N(0, \Sigma_{rv} \text{diagonal}) \quad (5)$$

Where

I_m = Latent personality variable indicator for r indicators for individual n

α_r = Unknown parameter of the indicator regarding the latent personality variable
 v_m = Error term in the indicator equation

The utility equation for personality is then given by

$$V_{in} = V_{in}(X_n, X_n^*; \beta) \quad i \in A(n), A(n) \text{ choice set}$$

Where

X_n = Observable variable, X_n^* = Latent variable, and β = Unknown parameters

The resulting utility functions are shown in equations 6 and 7.

$$V_{in} = \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_2 S_{1n}^* \quad (6)$$

$$U_{in} = V_{in} + \varepsilon_{in} \quad U_{in} = V_{in}(X_n, X_n^*; \beta) + \varepsilon_{in} \quad \varepsilon_n \sim D(0, \Sigma_\varepsilon)$$

$$U_{in} = \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_2 S_{1n}^* + \varepsilon_{in} \quad (7)$$

We can obtain the choice probability as Eq. 8.

$$P(Y_{in} = 1 | X_n, X_n^*; \beta) = \frac{e^{V_{in}}}{\sum_{j \in C_n} e^{V_{jn}}} \quad \forall_j \in A(n) \quad (8)$$

With

$$Y_{in} = \begin{cases} 1, & \text{if } U_{in} = \max_j \{U_{jn}\} \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

Where

Y_{in} = Indicator of choice, and U_{in} = Utility of alternative i for individual n

4. Model application

The proposed model was applied to a sample population of 218 people, 85% of whom are employed and 15% of whom are professors at the National University of Colombia at Medellin. Of the sample population, 53% are women; 43% men; 59% are over 35 years old; 52% had graduated level education; 48% have college degree, and

23% have incomes over \$2,000,000. The modal share of the sample was 24.8% auto; 45.9% bus; 6% taxi; 16.5% motorcycle; 3.2% walking and 3.6% Metro. The data were collected using revealed preference surveys.

4.1. Basic discrete choice model

This analysis corresponds to the results of a basic discrete choice model that does not include latent variables (estimated with the software BIOGEME); the model contains only alternative attributes and socio-economic characteristics of the elector. To assess the discrete model, auto, bus, taxi, motorcycle, walking and metro were used as alternatives.

4.2. Hybrid model with the personality variable

In this study, sex, age, and education were used as exogenous variables in the structural equations that determine the latent personality because these variables, especially sex and age, influence an individual's personality.

After applying equations (3) and (5) (using AMOS in SPSS software), the results shown in Table 1 were obtained.

Table 1 Parameters Personality traits

Factor		Personality traits	Parameter	t-test value
1	A	Reserved-Open	-0.944	(-2.018)
2	B	Concrete Thinking-Abstract Thinking	-1.422	(-2.618)
3	C	Emotional Instability-Emotional Stability	-3.993	(-9.224)
4	E	Submissive-Dominant	0.171	(0.375)
5	F	Prudent-Impulsive	-1.905	(-3.907)
6	G	Carefree-Scrupulous	-1.445	(-2.861)
7	H	Shy-Spontaneous	-2.269	(-5.765)
8	I	Rational-Emotional	-0.208	(-0.435)
9	L	Trusting-Suspicious	2.129	(4.611)
10	M	Practical-Dreamer	-0.105	(-0.211)
11	N	Single-Sly	-0.198	(-0.374)
12	O	Safe-Unsafe	3.463	(8.023)

Factor		Personality traits	Parameter	t-test value
13	Q1	Traditionalist-Innovative	0.696	(1.448)
14	Q2	Dependence on the group-Self-sufficient	0.360	(0.773)
15	Q3	Uninhibited-Controlled	-2.148	(-4.486)
16	Q4	Calm-Stressed	3.820	(9.661)

Overall, the sample population is a reserved community that poses concrete thinking and has a significant level of emotional instability. According to the parameters of the personality traits and their respective t-test, the community is shy, unsafe, and very stressed (see Table 1).

Table 2 shows the values of the parameter estimates (β) and their respective t-test, log likelihood and ρ^2 .

Table 2 Model Comparison

Variables	Basic MNL	Personality MNL
ASC1 auto	fixed	fixed
ASC2 bus	fixed	fixed
ASC3 taxi	-2.51	-1.79
	(-4.81)	(-3.23)
ASC4 motorcycle	fixed	fixed
ASC5 walk	-0.53	-0.595
	(-0.79)	(-0.85)
ASC6 metro	-2.44	-3.04
	(-2.5)	(-3.16)
(β_4) cost	-0.00016	-0.000166
	(-2.19)	(-2.03)
(β_8) walking time	-0.171	-0.0834
	(-4.36)	(-2.11)
(β_6) cost by income	0.000117	0.000102
	1.49	(+1.13)
(β_7) cost by sex	-0.00014	-0.000216
	(-1.59)	(-1.69)
(β_5) travel time	-0.0962	-0.0706
	(-5.5)	(-4.37)

Variables	Basic MNL	Personality MNL
(β_{per}) personality	-----	76.9
	-----	(+3.1)
$l(\beta)$	-110.439	-92.677
ρ^2	0.657	0.712

This model provides an excellent fit to the data, as all of the parameters have the correct signs (are conceptually valid) and are statistically appropriate. The most important variable is the latent personality variable, which is highly significant. Accordingly, for the sample population studied, personality is an important variable to take into account when modeling an individual's mode of transport.

Comparing the two models, the MNL Basic and the hybrid model with a personality variable (see Table 2), we found that the model considering the personality variable has better fit $l(\beta)$; $(-92.677 > -110.439)$ and a higher ρ^2 $(0.712 > 0.657)$. Furthermore, the latent personality variable is significant at the 95% confidence level (t-test is $3.1 > 1.96$); thus, the hybrid model is superior to the model that does not consider latent variables.

5. Conclusions

By integrating the latent personality variable into a discrete choice model, this study has presented a model that more accurately explains the decision process and thus has a smaller error term than that of the basic model. Econometric and psychometric discrete choice models have a better fit and are more explanatory than those that do not consider these factors. Thus, the models should be estimated using these two disciplines in a synergistic fashion. Hybrid discrete choice models that consider psychometric tests for the construction of latent variables provide more accurate results than models that do not do consider these tests.

The impacts of including the latent personality variable in a hybrid discrete choice model or using psychometric tests for the construction of latent variable indicators, which are then introduced into the hybrid model by sequential estimation, are unknown. This study has shown that a significant amount of research must be performed to incorporate the latent personality variable into the hybrid model by sequential estimation. More research should be conducted in this field to improve this type of model and thus advance the discrete choice model theory including personality variables.

References

- Ben-Akiva, M.E., Walker, J.L., Bernardino, A.T., Gopinath, D.A., Morikawa, T. and Polydoropoulou, A. (2002) Integration of choice and latent variable models. En H.S. Mahmassani (ed.), *In Perpetual Motion: Travel Behaviour Research Opportunities and Challenges*, 431-470. Pergamon, Amsterdam.
- Raveau, S., Alvarez-Daziano, R., Yáñez, M.F., Bolduc, D. and Ortúzar J.de D. (2010) Sequential and simultaneous estimation of hybrid discrete choice models: some new findings. 89th Annual Meeting of the Transportation Research Board. Washington, USA.
- Ashok, K., Dillon, W. and Yuan, S. (2002) Extending discrete choice models to incorporate attitudinal and other latent variables. *Journal of Marketing Research*, 39, 31-46.
- McFadden, D. (1986) The choice theory approach to marketing research. *Marketing Science* 5, 275-97.
- Morikawa, T., Sasaki, K., (1998) Discrete choice models with latent variables using subjective data. In: Ortúzar, J. de D., Hensher, D.A., Jara-Díaz, S. (Eds.), *Travel Behavior. Research: Updating the State of Play*. Pergamon, Oxford, pp. 435-455.
- Morikawa, T., Ben-Akiva, M., McFadden, D., (2002) Discrete Choice Models Incorporating Revealed Preferences and Psychometric Data. *Econometric Models in Marketing Advances in Econometrics: A Research Annual*, vol. 16. Elsevier Science Ltd.
- Pendleton, L.H., Shonkwiler, J.S., (2001) Valuing bundled attributes: a latent characteristics approach. *Land Economics* 77 (1), 118-129.
- Vredin, Johansson, M., Heldt, T. and Johansson, P. (2005) Latent variables in a travel mode choice model: Attitudinal and behavioural

- indicator variables. Working Paper, Department of Economics, Uppsala University.
- Yáñez, M.F., Raveau, S., Ortúzar, J. de D. (2010) Inclusion of latent variables in Mixed logit models: Modeling and forecasting. *Transportation Research Part A* 44,744-753.
- Walker, J. (2001) *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures and Latent Variables*. PhD. Thesis, MIT.
- Domencich, T. and McFadden, D., (1975) *Urban Travel Demand: A Behavioral Analysis*, Ed. North-Holland, New York.
- Williams, H. and Ortúzar, J., (1982) Behavioral theories of dispersion and the misspecification of travel demand models, *Transportation Research*, 16B (3), 167-219.
- Cattell, R. and Eysenck, H. (1967) *The Biological Basis of Personality*.
- Cattell, R. B., Eber, H. W. and Tatsuoka, M. M., (1970) *Handbook for the sixteen personality factor questionnaire (16PF)*, Institute for Personality and Ability Testing, Champaign, IL.